



OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/044,447

DATE: 08/14/2002

TIME: 15:17:44

Input Set : N:\CrF3\RULE60\10044447.raw

Output Set: N:\CRF3\08142002\J044447.raw

1 <110> APPLICANT: Petrini, John H.J.
 2 Morgan, William Franklin
 3 Maser, Richard Scott
 4 Carney, James Patrick
 5 <120> TITLE OF INVENTION: DNA Encoding A DNA Repair Protein
 6 <130> FILE REFERENCE: 800.019US1
 7 <140> CURRENT APPLICATION NUMBER: 10/044,447
 8 <141> CURRENT FILING DATE: 2002-01-10
 11 <150> PRIOR APPLICATION NUMBER: US/09/067,641
 12 <151> PRIOR FILING DATE: 1998-04-27
 14 <160> NUMBER OF SEQ ID NOS: 24
 15 <170> SOFTWARE: FastSEQ for Windows Version 3.0
 17 <210> SEQ ID NO: 1
 18 <211> LENGTH: 4403
 19 <212> TYPE: DNA
 20 <213> ORGANISM: Homo sapiens
 21 <400> SEQUENCE: 1

ENTERED

22	ttcggcacga ggcgcggttg cacgtcggcc ccagccctga ggagccggac cgatgtggaa	60
23	actgctgcc gccgcgggccc cggcaggagg agaaccatac agacttttga ctggcggttg	120
24	gtacgttggtt ggaaggaaaa actgtgccat tctaattgaa aatgatcagt cgatcagccc	180
25	aaatcatgct gtgttaactg ctaacttttc tgtaaccaac ctgagtcaaa cagatgaaat	240
26	ccctgtattg acattaaaaag ataattctaa gtatggtacc tttgttaatg aggaaaaaat	300
27	gcagaatggc ttttcccgaa ctttgaagtc gggggatggt attacttttg gagtgtttgg	360
28	aagtaaatte agaataagat atgagccttt ggttgcatgc tcttcttggt tagatgtctc	420
29	tgggaaaact gctttaaatc aagctatat gcaacttgga ggatttactg taaacaattg	480
30	gacagaagaa tgcactcacc ttgtcatggt atcagtgaat gttaccatta aaacaatatg	540
31	tgcactcatt tgtggacgtc caattgtaaa gccagaatat tttactgaat tcttgaaagc	600
32	agttcagtc aagaagcagc ctccacaaat tgaaagtttt taccacacct ttgatgaacc	660
33	atctattgga agtaaaaatg ttgatctgtc aggacggcag gaaagaaaac aaatcttcaa	720
34	agggaaaaca tttatatatt tgaatgccaa acagcataag aaattgagtt ccgcagttgt	780
35	ctttggaggt ggggaagcta ggttgataac agaagagaat gaagaagaac ataatttctt	840
36	tttggctccg ggaacgtgtg ttgttgatac aggaataaca aactcacaga ccttaattcc	900
37	tgactgtcag aagaaatgga ttcagtcaat aatggatatg ctccaaaggc aagggtcttag	960
38	acctattcct gaagcagaaa ttggattggc ggtgattttc atgactacaa agaattactg	1020
39	tgatcctcag ggccatccca gtacaggatt aaagacaaca actccaggac caagcctttc	1080
40	acaaggcgtg tcagttgatg aaaaactaat gccaaagccc ccagtgaaca ctacaacata	1140
41	cgtagctgac acagaatcag agcaagcaga tacatgggat ttgagtgaat ggccaaaaga	1200
42	aatcaaagtc tccaaaatg aacaaaaatt cagaatgctt tcacaagacy caccactgt	1260
43	aaaggagtc tgcaaaaaca gctctaata taatagtatg gtatcaaata ctttggctaa	1320
44	gatgagaatc ccaaactatc agctttcacc aactaaattg ccaagtataa ataaaagtaa	1380
45	agataggggt tctcagcagc agcagaccaa ctccatcaga aactactttc agccgtctac	1440
46	caaaaaaagg gaaagggatg aagaaaatca agaaatgtct tcatgcaaat cagcaagaat	1500
47	agaaacgtct tgttctcttt tagaacaac acaacctgct acacctcat tgtggaaaaa	1560

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/044,447

DATE: 08/14/2002

TIME: 15:17:44

Input Set : N:\Crif3\RULE60\10044447.raw

Output Set: N:\CRF3\08142002\J044447.raw

```

48 taaggagcag catctatctg agaatgagcc tgtggacaca aactcagaca ataacttatt 1620
49 tacagataca gatttaaaat ctattgtgaa aaattctgcc agtaaatctc atgctgcaga 1680
50 aaagctaaga tcaaataaaa aaagggaat ggatgatgtg gccatagaag atgaagtatt 1740
51 ggaacagtta ttcaaggaca caaaaccaga gttagaaatt gatgtgaaag ttcaaaaaca 1800
52 ggaggaagat gtcaatgtta gaaaaaggcc aaggatggat atagaaacaa atgacacttt 1860
53 cagtgatgaa gcagtaccag aaagtagcaa aatatctcaa gaaaatgaaa ttgggaagaa 1920
54 acgtgaactc aggaagact cactatggtc agctaaagaa atatctaaca atgacaaact 1980
55 tcaggatgat agtgagatgc ttccaaaaaa gctgttattg actgaattta gatcactggt 2040
56 gattaaaaac tctacttcca gaaatccgtc tggcataaat gatgattatg gtcaactaaa 2100
57 aaatttcaag aaattcaaaa aggtcacata tcctggagca ggaaaacttc cacacatcat 2160
58 tggaggatca gatctaatac ctcatcatgc tcgaaagaat acagaactag aagagtggct 2220
59 aaggcaggaa atggaggtac aaaatcaaca tgcaaaagaa gagtctcttg ctgatgatct 2280
60 ttttagatac aatccttatt taaaaaggag aagataactg aggattttta aaagaagcca 2340
61 tggaaaaact tcctagtaag catctacttc aggccaccaa ggttatatga atatatagt 2400
62 tatagaagcg atttaagtta caatgtttta tggcctaaat ttattaaata aaatgcacaa 2460
63 aactttgatt cttttgtatg taacaattgt ttgtyctgtt ttcaggcttt gtcattgcat 2520
64 ctttttttca tttttaaatg tgttttgttt attaaatagt taatatagtc acagttcaaa 2580
65 attctaaatr tacgtaaggt aaaggactaa agtcaccctt ccaccattgt cctagctact 2640
66 tattttttaa taatttccca cacaaatgat agcataacat atgcagtgtt ctacaccttg 2700
67 ctttttttact tagtaagatt aaaaattata ggaatatcaa tataatgttt ttaatatatt 2760
68 ttcttttcca ttatgctgta gtcttaccta aactctggtg atccaaacaa aatggcttca 2820
69 gtggtgcaga tgtcacctac atgttattct agtactagaa actgaagacc atgtggagac 2880
70 ttcatcaaac atgggtttag ttttcaccag aatggaaaga cctgtacccc ttttgggtgg 2940
71 tcttactgag ctgggtgggt gtctgttttg agcttattta gagtccctagt tttcctactt 3000
72 ataaagtaga aatggtgaga ttgttttctt tttctacckt aaaggagat ggtaagaaac 3060
73 aatgaatgtc ttttttcaaa ctttattgac aagtgtttt caagtctgtg ttcaaaaata 3120
74 tattcatgta cctgtgatcc agcaagaagg gagttccagt caagagtcac tacaactgat 3180
75 tagttgttta gagaatgaga aatggaacag tgaggaatgg aggccatatt tccatgactt 3240
76 cccttgtaaa cagaagcaac agaagggaca agaggctggc ctctacatca ctctcacctt 3300
77 ccaaactctg tggaagtgc tctacttgcc agaaccaaat taacttactt ccaagttctg 3360
78 gctgcttgca ggtggaactc cagctgcaag ggagttaggg aaatgaaggt ctttttttaa 3420
79 aagcttctca gccttcttag ggaacagaaa ttgggtgagc caatctgcaa tttctactac 3480
80 aggcattgag accagttaga ttattgaaat attatagaga gttatgaaca cttaaattat 3540
81 gatagtggta tgacattgga tagaacatgg gatactttag aagtagaatt gacagggcat 3600
82 attagtgtat gaaatggagt catttgagtc tyttaatagc catgtatcat aattaccaag 3660
83 tgaagctggt ggaacatatg gtctccattt tacagttaag gaatataatg gacagattaa 3720
84 tattgttytc tgtcatgccc acaatccctt tctaaggaa actgccctac tatagcagtt 3780
85 tttatatatt tcaatttatg aatataatga atgaggagtt ctggtacctc ctgtctttac 3840
86 aaatattggg tgttgtccag tatttttccc tttttaaccc ttcccaattc ggggtgtgtg 3900
87 gtggatgttt ccatttggggt tttaatgtgt atatccctga tagctataat tgggtcatag 3960
88 aaattcttta tacattctag atgcaagtc cttgycggt atacgtattg agatattaca 4020
89 cctagtctgt ggcttgactg tttcttttat gtcttttgat gaatagaagt tttaaatttt 4080
90 gacaaggtca aatttatttt tttcttttgt ttgatatttt ttctctccaa tttaaaccca 4140
91 agatttcaga tattctgctc tattatataa actttatatt tttatatttg tgatctacct 4200
92 tgaattgata tgtatgttgt gaattatgga tcagggttct tttttcccc catacaagta 4260
93 tccagtcat gtaacactgt ttattgaaag aattatcctt tctcattaa attaccttgc 4320
94 caattagtaa aaaatcaatt aaccatrmr mmmrrrggat ccactagttc tagagcggcc 4380
95 gccaccgcgg tggagctcca gct 4403
97 <210> SEQ ID NO: 2

```

RAW SEQUENCE LISTING

DATE: 08/14/2002

PATENT APPLICATION: US/10/044,447

TIME: 15:17:44

Input Set : N:\Crif3\RULE60\10044447.raw

Output Set: N:\CRF3\08142002\J044447.raw

98 <211> LENGTH: 754

99 <212> TYPE: PRT

100 <213> ORGANISM: Homo sapiens

101 <400> SEQUENCE: 2

```

102      Met Trp Lys Leu Leu Pro Ala Ala Gly Pro Ala Gly Gly Glu Pro Tyr
103      1          5          10          15
104      Arg Leu Leu Thr Gly Val Glu Tyr Val Val Gly Arg Lys Asn Cys Ala
105      20          25          30
106      Ile Leu Ile Glu Asn Asp Gln Ser Ile Ser Arg Asn His Ala Val Leu
107      35          40          45
108      Thr Ala Asn Phe Ser Val Thr Asn Leu Ser Gln Thr Asp Glu Ile Pro
109      50          55          60
110      Val Leu Thr Leu Lys Asp Asn Ser Lys Tyr Gly Thr Phe Val Asn Glu
111      65          70          75          80
112      Glu Lys Met Gln Asn Gly Phe Ser Arg Thr Leu Lys Ser Gly Asp Gly
113      85          90          95
114      Ile Thr Phe Gly Val Phe Gly Ser Lys Phe Arg Ile Glu Tyr Glu Pro
115      100          105          110
116      Leu Val Ala Cys Ser Ser Cys Leu Asp Val Ser Gly Lys Thr Ala Leu
117      115          120          125
118      Asn Gln Ala Ile Leu Gln Leu Gly Gly Phe Thr Val Asn Asn Trp Thr
119      130          135          140
120      Glu Glu Cys Thr His Leu Val Met Val Ser Val Lys Val Thr Ile Lys
121      145          150          155          160
122      Thr Ile Cys Ala Leu Ile Cys Gly Arg Pro Ile Val Lys Pro Glu Tyr
123      165          170          175
124      Phe Thr Glu Phe Leu Lys Ala Val Gln Ser Lys Lys Gln Pro Pro Gln
125      180          185          190
126      Ile Glu Ser Phe Tyr Pro Pro Leu Asp Glu Pro Ser Ile Gly Ser Lys
127      195          200          205
128      Asn Val Asp Leu Ser Gly Arg Gln Glu Arg Lys Gln Ile Phe Lys Gly
129      210          215          220
130      Lys Thr Phe Ile Phe Leu Asn Ala Lys Gln His Lys Lys Leu Ser Ser
131      225          230          235          240
132      Ala Val Val Phe Gly Gly Gly Glu Ala Arg Leu Ile Thr Glu Glu Asn
133      245          250          255
134      Glu Glu Glu His Asn Phe Phe Leu Ala Pro Gly Thr Cys Val Val Asp
135      260          265          270
136      Thr Gly Ile Thr Asn Ser Gln Thr Leu Ile Pro Asp Cys Gln Lys Lys
137      275          280          285
138      Trp Ile Gln Ser Ile Met Asp Met Leu Gln Arg Gln Gly Leu Arg Pro
139      290          295          300
140      Ile Pro Glu Ala Glu Ile Gly Leu Ala Val Ile Phe Met Thr Thr Lys
141      305          310          315          320
142      Asn Tyr Cys Asp Pro Gln Gly His Pro Ser Thr Gly Leu Lys Thr Thr
143      325          330          335
144      Thr Pro Gly Pro Ser Leu Ser Gln Gly Val Ser Val Asp Glu Lys Leu
145      340          345          350
146      Met Pro Ser Ala Pro Val Asn Thr Thr Thr Tyr Val Ala Asp Thr Glu

```

TIME: 15:17:44

Output Set: N:\CRF3\08142002\J044447.raw

147				355				360					365			
148	Ser	Glu	Gln	Ala	Asp	Thr	Trp	Asp	Leu	Ser	Glu	Arg	Pro	Lys	Glu	Ile
149		370					375					380				
150	Lys	Val	Ser	Lys	Met	Glu	Gln	Lys	Phe	Arg	Met	Leu	Ser	Gln	Asp	Ala
151	385					390					395					400
152	Pro	Thr	Val	Lys	Glu	Ser	Cys	Lys	Thr	Ser	Ser	Asn	Asn	Asn	Ser	Met
153					405					410					415	
154	Val	Ser	Asn	Thr	Leu	Ala	Lys	Met	Arg	Ile	Pro	Asn	Tyr	Gln	Leu	Ser
155				420					425					430		
156	Pro	Thr	Lys	Leu	Pro	Ser	Ile	Asn	Lys	Ser	Lys	Asp	Arg	Ala	Ser	Gln
157			435					440					445			
158	Gln	Gln	Gln	Thr	Asn	Ser	Ile	Arg	Asn	Tyr	Phe	Gln	Pro	Ser	Thr	Lys
159		450					455					460				
160	Lys	Arg	Glu	Arg	Asp	Glu	Asn	Gln	Glu	Met	Ser	Ser	Cys	Lys	Ser	
161	465					470				475					480	
162	Ala	Arg	Ile	Glu	Thr	Ser	Cys	Ser	Leu	Leu	Glu	Gln	Thr	Gln	Pro	Ala
163					485					490					495	
164	Thr	Pro	Ser	Leu	Trp	Lys	Asn	Lys	Glu	Gln	His	Leu	Ser	Glu	Asn	Glu
165				500					505					510		
166	Pro	Val	Asp	Thr	Asn	Ser	Asp	Asn	Asn	Leu	Phe	Thr	Asp	Thr	Asp	Leu
167			515					520					525			
168	Lys	Ser	Ile	Val	Lys	Asn	Ser	Ala	Ser	Lys	Ser	His	Ala	Ala	Glu	Lys
169		530					535					540				
170	Leu	Arg	Ser	Asn	Lys	Lys	Arg	Glu	Met	Asp	Asp	Val	Ala	Ile	Glu	Asp
171	545					550					555					560
172	Glu	Val	Leu	Glu	Gln	Leu	Phe	Lys	Asp	Thr	Lys	Pro	Glu	Leu	Glu	Ile
173					565					570					575	
174	Asp	Val	Lys	Val	Gln	Lys	Gln	Glu	Glu	Asp	Val	Asn	Val	Arg	Lys	Arg
175				580					585					590		
176	Pro	Arg	Met	Asp	Ile	Glu	Thr	Asn	Asp	Thr	Phe	Ser	Asp	Glu	Ala	Val
177			595					600					605			
178	Pro	Glu	Ser	Ser	Lys	Ile	Ser	Gln	Glu	Asn	Glu	Ile	Gly	Lys	Lys	Arg
179		610					615					620				
180	Glu	Leu	Lys	Glu	Asp	Ser	Leu	Trp	Ser	Ala	Lys	Glu	Ile	Ser	Asn	Asn
181	625					630					635					640
182	Asp	Lys	Leu	Gln	Asp	Asp	Ser	Glu	Met	Leu	Pro	Lys	Lys	Leu	Leu	Leu
183					645					650					655	
184	Thr	Glu	Phe	Arg	Ser	Leu	Val	Ile	Lys	Asn	Ser	Thr	Ser	Arg	Asn	Pro
185				660					665					670		
186	Ser	Gly	Ile	Asn	Asp	Asp	Tyr	Gly	Gln	Leu	Lys	Asn	Phe	Lys	Lys	Phe
187			675													

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/044,447

DATE: 08/14/2002

TIME: 15:17:44

Input Set : N:\Crif3\RULE60\10044447.raw

Output Set: N:\CRF3\08142002\J044447.raw

```

196      Arg Arg
198 <210> SEQ ID NO: 3
199 <211> LENGTH: 87
200 <212> TYPE: PRT
201 <213> ORGANISM: Homo sapiens
202 <220> FEATURE:
203 <221> NAME/KEY: UNSURE
204 <222> LOCATION: (48)...(48)
205 <223> OTHER INFORMATION: Unsure
206 <400> SEQUENCE: 3
207      Tyr Val Val Gly Arg Lys Asn Cys Ala Ile Leu Ile Glu Asn Asp Gln
208          1             5             10             15
209      Ser Ile Ser Arg Asn His Ala Val Leu Thr Ala Asn Phe Ser Val Thr
210          20             25             30
W--> 211      Asn Leu Ser Gln Thr Asp Glu Ile Pro Val Leu Thr Leu Lys Asn Xaa
212          35             40             45
213      Lys Tyr Gly Thr Phe Val Asn Glu Glu Lys Met Gln Asn Gly Phe Ser
214          50             55             60
215      Arg Thr Leu Lys Ser Val Asp Gly Ile Thr Phe Gly Val Phe Gly Ser
216          65             70             75             80
217      Lys Phe Arg Ile Glu Tyr Glu
218          85
220 <210> SEQ ID NO: 4
221 <211> LENGTH: 87
222 <212> TYPE: PRT
223 <213> ORGANISM: Homo sapiens
224 <400> SEQUENCE: 4
225      Tyr Ser Ile Gly Arg Ser Ser Lys Asn Pro Leu Ile Ile Lys Asn Asp
226          1             5             10             15
227      Lys Ser Ile Ser Arg Gln His Ile Thr Phe Lys Trp Glu Ile Asn Asn
228          20             25             30
229      Ser Ser Asp Leu Lys His Ser Ser Lys Cys Leu Val Asn Lys Gly Lys
230          35             40             45
231      Leu Thr Ser Leu Asn Lys Lys Phe Met Lys Val Gly Glu Thr Phe Thr
232          50             55             60
233      Ile Asn Ala Ser Cys Val Leu Lys Ser Thr Ile Glu Leu Gly Thr Thr
234          65             70             75             80
235      Pro Ile Arg Ile Glu Phe Glu
236          85
238 <210> SEQ ID NO: 5
239 <211> LENGTH: 13
240 <212> TYPE: PRT
241 <213> ORGANISM: Homo sapiens
242 <400> SEQUENCE: 5
243      Asn Pro Ser Gly Leu Asn Asp Asp Tyr Gly Gln Leu Lys
244          1             5             10
246 <210> SEQ ID NO: 6
247 <211> LENGTH: 680
248 <212> TYPE: PRT

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/044,447

DATE: 08/14/2002

TIME: 15:17:45

Input Set : N:\Crf3\RULE60\10044447.raw

Output Set: N:\CRF3\08142002\J044447.raw

L:211 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3